

INFORMATION PROCESSOR COLLECTING AND MANAGING LOG DATA

Publication number: JP10293704 (A)

Also published as:

Publication date: 1998-11-04

JP3778652 (B2)

Inventor(s): FUJINO SHUJI; MORIKAWA TOSHIYOSHI; URANO AKIHIRO;
NAKANO HIDENORI; MORITA SHINJI; YAMADA MITSUGI;
NIMURA YOSHITAKA *

US8173418 (B1)

Applicant(s): HITACHI LTD *

Classification:

- International: G06F11/00; G06F11/34; (IPC1-7): G06F11/34

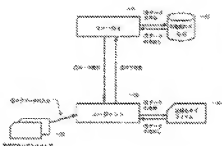
- European: G06F11/3474; H04L12/24A4; H04L12/24D

Application number: JP19970101210 19970418

Priority number(s): JP19970101210 19970418

Abstract of JP 10293704 (A)

PROBLEM TO BE SOLVED: To manage log data based on a common data format and also to manage log data based on the time of a manager in a system in which a manager collects log data from plural agents through a network. SOLUTION: A manager 10 distributes various rules to an agent 20. The agent 20 inputs log data from a monitored object log file group 30 according to the rules, normalizes it, adds the correction time of a log output time and stores it in a normalization log file 40. The agent 20 fetches normalization log data from the file 40 and transfers it to the manager 10 according to a request from the manager 10. The manager 10 stores collected normalization log data in a normalization log database 50 in order of the correction time.



.....
Data supplied from the **espacenet** database — Worldwide

特開平10-293704

(43)公開日 平成10年(1998)11月4日

(51)Int.Cl.⁶

識別記号

F i

G 0 6 F 11/34

G 0 6 F 11/34

B

審査請求 未請求 請求項の数5 O L (全 24 頁)

(21)出願番号 特願平9-101210

(22)出願日 平成9年(1997)4月18日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 藤野 修司

神奈川県横浜市戸塚区戸塚町5030番地 株

式会社日立製作所ソフトウェア開発本部内

(72)発明者 森川 寿義

東京都千代田区大手町二丁目6番2号 株

式会社日立情報ネットワーク内

(72)発明者 浦野 明裕

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74)代理人 弁理士 高橋 明夫

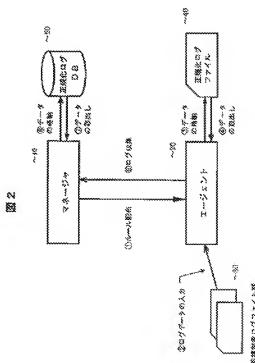
最終頁に続く

(54)【発明の名称】 ログデータの収集と管理をする情報処理装置

(57)【要約】

【課題】 マネージャがネットワークを介して複数のエージェントからログデータを収集するシステムにおいて、共通的なデータ形式に基づいてログデータを管理する。またマネージャの時刻に基づいたログデータを管理する。

【解決手段】 マネージャ10は、各種ルールをエージェント20に配布する。エージェント20は、このルールに従って監視対象ログファイル群30からログデータを入力し、正規化し、ログ出力時刻の補正時刻を付加して正規化ログファイル40に格納する。マネージャ10からの要求に従ってエージェント20は、正規化ログファイル40から正規化ログデータを取り出してマネージャ10へ転送する。マネージャ10は、収集した正規化ログデータを補正時刻の順に正規化ログデータベース50に格納する。



【特許請求の範囲】

【請求項1】 監視の対象とするログファイル中のログデータからあらかじめ定義されたデータ項目に対応する値を切り出して規定されたデータ項目の値を配列する正規化されたログデータを作成して蓄積する手段と、蓄積された正規化ログデータをネットワークを介してマネージャの機能を実行する情報処理装置へ送信する手段とをエージェントの処理手段として有することを特徴とするログデータの収集をする情報処理装置。

【請求項2】 コンピュータ読み取り可能な記憶媒体上に実体化され、ログデータを収集するエージェント機能を有するコンピュータプログラムであって、該プログラムは以下のステップを含む：

(a) 監視の対象とするログファイル中のログデータからあらかじめ定義されたデータ項目に対応する値を切り出して規定されたデータ項目の値を配列する正規化されたログデータを作成し、(b) 蓄積された正規化ログデータをネットワークを介してマネージャの機能を実行するコンピュータへ送信する。

【請求項3】 ネットワークを介してエージェントの機能を実行する情報処理装置からあらかじめ定義された共通のデータ形式に従って正規化されたログデータであってマネージャの基準とする時刻に基づいて補正されたログ出力時刻を有する正規化ログデータを受信する手段と、該正規化ログデータを補正時刻の順にデータベースに蓄積する手段とをマネージャの処理手段として有することを特徴とするログデータの収集と管理をする情報処理装置。

【請求項4】 該補正時刻とマネージャの現在時刻との差分の時間が所定の保存期間を超過している正規化ログデータを該データベースから削除する手段をさらに設けることを特徴とする請求項3記載のログデータの収集と管理をする情報処理装置。

【請求項5】 コンピュータ読み取り可能な記憶媒体上に実体化され、ログデータの収集と管理をするマネージャ機能を有するコンピュータプログラムであって、該プログラムは以下のステップを含む：

(a) ネットワークを介してエージェントの機能を実行するコンピュータからあらかじめ定義された共通のデータ形式に従って正規化されたログデータであってマネージャの基準とする時刻に基づいて補正されたログ出力時刻を有する正規化ログデータを受信し、(b) 該正規化ログデータを補正時刻の順にデータベースに蓄積する。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、マネージャがネットワークを介してエージェントからログデータを収集するシステムに係わり、特にマネージャがシステム内に存在するログデータを共通的なログデータ形式に基づいて管理するログデータの収集と管理の方式に関する。

【0002】

【従来の技術】 情報処理装置で実行されるオペレーティングシステム（OS）やアプリケーションプログラムは、各種のログ情報を出力する。出力されたログ情報は、各種のログ情報として蓄積される。例えば特開平5-250229号公報は、複数のコンピュータからのログデータ収集において、ログデータ中のエラーコードを検出することにより、エラー状態のコンピュータからのログデータを優先的に送信するログデータ収集方式を開示する。また特開平5-28008号公報は、情報処理システムが障害ログを収集するときに障害手段に附えられたログ情報の個数が一定数に達したことを検出し、ログ登録の抑止を行うことにより重要な障害情報の消失を防ぐログ情報収集方式を開示する。また特開平5-11029号公報は、下位の設備端末からのデータが採取された時刻をデータに付与して上位の制御装置に送ることにより、複数の端末からの各データの時系列的な関係が明なわることがないようにするデータ収集方式を開示する。

【0003】

【発明が解決しようとする課題】 ネットワークを介して複数のコンピュータが接続され、コンピュータが他のコンピュータと通信しながら処理を進める分散処理システムにおいて、一人のユーザは複数の制御範囲に亘るコンピュータやファイルにアクセスし得る。従ってログデータを解析することによってコンピュータやファイルへの不正なアクセスを検出するためには、個々のコンピュータが出力するログデータを中央のコンピュータに集約し、データベースに蓄積する必要がある。しかしながら各種のシステムプログラムやアプリケーションプログラムが出力するログデータは、各々そのデータ形式が異なっているため、単に個々のコンピュータプログラムが出力するログデータを収集して集約するだけではログデータの解析が困難である。また個々のコンピュータが保有する時刻がすべてのコンピュータに亘って一致しているとは限らないので、一般に個々のログデータに付与されている時刻にはずれが生じており、集約されたログデータを正しい時刻の順序に従って配列することが困難である。

【0004】 本発明は、上記の事情に鑑みてなされたものであり、その目的とするところは、共通的なデータ形式に正規化されたログデータを管理することにある。

【0005】 本発明の他の目的は、各サイトのコンピュータからログデータを収集する中央のコンピュータの時刻を基準とするログデータを管理することにある。

【0006】

【課題を解決するための手段】 上記の目的を達成するため、本発明は、監視の対象とするログファイル中のログデータからあらかじめ定義されたデータ項目に対応する値を切り出して規定されたデータ項目の値を配列する正規化されたログデータを作成して蓄積する手段と、蓄積

された正規化ログデータをネットワークを介してマネージャの機能を実行するコンピュータへ送信する手段とをエージェントの処理手段として有するログデータの収集をするコンピュータを特徴とする。

【0007】また本発明は、ネットワークを介してエージェントの機能を実行するコンピュータからあらかじめ定義された共通のデータ形式に従って正規化されたログデータであってマネージャの基準とする時刻に基づいて補正されたログ出力時刻を有する正規化ログデータを受信する手段と、該正規化ログデータを補正時刻の順にデータベースに蓄積する手段とをマネージャの処理手段として有するログデータの収集と管理をするコンピュータを特徴とする。

【0008】

【発明の実施の形態】以下、本発明の実施形態について図面に基づいて詳細に説明する。

【0009】図1は、本発明を適用する通信ネットワークの一例を示すシステム構成図である。ネットワークは複数のLAN1、3、4及びWAN（ワイド・エリア・ネットワーク）2に結合されるものである。マネージャ10は、計算機の主記憶装置に格納され、OSの下で実行するアプリケーションプログラム（AP）である。エージェント20は、サーバ等の計算機の主記憶装置に格納され、OSの下で実行するAPである。マネージャ10は、LAN1、3、4又はWAN2を介してエージェント20-1、20-2、20-3、20-4と通信することが可能である。図に示すように、エージェント20を実行する計算機はエージェント20と並行して他のAPを実行することが可能である。マネージャ10を実行する計算機もマネージャ10と並行して他のAPを実行することが可能であるが、充分なCPU性能と資源を確保するためには、専用の計算機を割り当てるのが望ましい。60はこのネットワークシステムを管理するネットワーク管理システムであり、ネットワーク管理プログラム及び図示しないネットワーク監視端末から構成されるネットワーク管理システム60は、ネットワーク資源についての情報をマネージャ10に提供する。

【0010】図2は、マネージャとエージェントが行う処理動作の概略を示す図である。マネージャ10は、定義された各種ルールをエージェント20へ配布する(④)。エージェント20は、受信したルールを登録し、そのルールに従って監視対象ログファイル群30からログデータを入手し(⑤)、ログデータを正規化した後、正規化ログファイル40に格納する(⑥)。このとき正規化ログデータ中のログ出力時刻をマネージャ10の時刻を基準とする補正時刻によって補正する。エージェント20

は、マネージャ10からログ収集要求があったとき正規化ログファイル40から正規化されたログデータを取り出し(⑦)、マネージャ10へ転送する(⑧)。マネージャ10は、収集した正規化ログデータを補正時刻の順に正規化ログデータベース50に格納する(⑨)。またマネージャ10は、必要に応じて正規化ログデータベース50から任意の正規化ログデータを抽出し、その解析を行う。

【0011】図3は、マネージャ10の構成を示す機能ブロック図である。ルール110は定義された各種ルールであり、記憶装置に格納される。正規化ログデータベース50は、収集された正規化ログデータを格納するデータベースであり、記憶装置に格納される。ネットワーク管理システム60は、ネットワークを介してマネージャ10を格納する計算機と接続される他の計算機およびネットワークを監視する端末装置から構成されるシステムであり、マネージャ10と同じ計算機で実行するネットワーク管理プログラムは、ネットワークの構成要素である各通信回線、ルータ、中継機、各種計算機、計算機のプログラム等の動作状態（接続中／接続断、動作中／停止状態など）、各ネットワーク構成要素の所在場所などを管理する。以下マネージャ10を構成する各機能モジュールの機能の概略について述べる。

【0012】(1)ルール定義100

運用者が各種ルールを定義するためのプログラムツールであり、図示しない入力装置及び表示装置を介してユーザが容易にルールを設定できるようなGUI（グラフィカル・ユーザ・インターフェース）を提供する。

【0013】(2)スケジューラ101

ルール配布及びログ収集を実行する頻度をそれぞれルール配布106及びログ収集107に知らせるプログラムである。頻度の例として、例えば毎日午後5時、毎週土曜日の午後3時15分などのように実行開始を指示する。

【0014】(3)プロセス管理102

マネージャ10の100～109を付与する各機能モジュールの起動／停止を制御するプログラムである。

【0015】(4)ログ解析103

データベース管理109を介して正規化ログデータベース50から正規化ログデータ群を抽出し、所定の解析を行うプログラムである。以下ログ対象がログインである場合のログデータ解析の例を表1に示す。ログ対象及び正規化項目については後述する。

【0016】

【表1】

ログデータ解析の例(表1)

ログ解析項目	使用する正規化項目
規定のログイン地域でない所からログイン成功した	ユーザ名 接続元ホスト名 接続元IPアドレス
規定の時間外にログインして失敗した	ユーザ名 開始時刻
同一ユーザが別々の地域から同時にログインしている	ユーザ名 接続元ホスト名
一定期間に規定回数以上のログイン回数がありログインに失敗した	ログ事象 ログ事象結果 ユーザ名 補正時刻
規定回数以上、ユーザ名とパスワードの組み合わせを失敗したユーザ変更を行った	ログ事象 ログ事象結果 ユーザ名 変更後のユーザ名 補正時刻

【0017】(5) 構成管理104

対象とするエージェント20の一覧を管理するプログラムである。またネットワーク管理システム60に問い合わせを行ってエージェント20が動作する計算機の動作有無やpingの応答時間等の情報を取得してルール配布106又はログ収集107に渡す。

【0018】(6) ルール管理105

ルール定義100から定義された各種ルールを受け取ってルール110に格納する。またルール110からルールを読み出してルール配布106に渡す。

【0019】(7) ルール配布106

スケジューラ101の指示に従い、ルール管理105から各種ルールを取得してエージェント20へ配布するプログラムである。

【0020】(8) ログ収集107

スケジューラ101の指示に従い、エージェント20から正規化ログファイルを収集するプログラムである。

【0021】(9) データ通信108

ルール配布106及びログ収集107がエージェント20と通信するときに通信制御を行うプログラムである。

【0022】(10) データベース管理109

ログ収集107がエージェント20から収集した正規化ログデータ群(正規化ログファイル)を正規化ログデータベース50に格納し、ログ解析103からの要求によって正規化ログデータを検索し抽出する。また所定の保存期間を過ぎた正規化ログデータを正規化ログデータベース50から削除し、未使用の記憶領域を生み出す。

【0023】図4は、エージェント20の構成を示す機能ブロック図である。ルール205は配布を受けた各種ルールであり、記憶装置に格納される。監視対象ログファイル群30は、監視対象とするログファイルであり、

記憶装置に格納される。正規化ログファイル40は、正規化ログデータを格納するファイルであり、記憶装置に格納される。以下エージェント20を構成する各機能モジュールの機能の概略について述べる。

【0024】(1) データ通信200

データ通信200は、ルール管理203及びログファイル管理204がマネージャ10と通信するときに通信制御を行うプログラムである。

【0025】(2) プロセス管理201

エージェント20の200〜204を付す各機能モジュールの起動/停止を制御するプログラムである。

【0026】(3) ログ入力202

ルール管理203からログファイル監視ルールやフォーマットルール等を取得し、監視対象ログファイル群30から入力したログデータを正規化した後、正規化ログデータをログファイル管理204へ渡すプログラムである。

【0027】(4) ルール管理203

マネージャ10から配布された各種ルールをルール205として記憶装置に格納し、ログ入力202又はログファイル管理204の要求に応じてルールを提供するプログラムである。

【0028】(5) ログファイル管理204

ルール管理203からフィルタリングルールを取得し、ログ入力202から受け取った正規化ログデータをフィルタリングし、補正時刻を付加して正規化ログファイル40に格納する。マネージャ10のログ収集107から正規化ログファイル40の取集要求を受信したとき、正規化ログファイル40をマネージャ10へ返送する。

【0029】図5〜図14は、正規化ログデータの構造の一例を示す図である。

【0030】図5は、正規化ログファイル40に格納される正規化ログデータ300の概略構成を示す図である。正規化ログデータ300は、共通情報クラス301と必要に応じて追加されるユーザ情報クラス302、サービス情報クラス303、アドレス情報クラス304、ファイル情報クラス305、トラフィック情報クラス306、個別情報クラス307等から構成される。共通情報クラス301は、すべての正規化ログデータ300に必要な情報クラスであり、残りの情報クラスは出力されたログデータに応じて選択されるものである。

【0031】図6は、共通情報クラス301のデータ構成を示す図である。正規化バージョン310は、正規化のバージョンを示す番号である。ログ種別はログ事象311、ログ事象結果312、ログ出力プログラム313、データ格納クラス314及びログファイル名315を含む。ログ出力プログラム313はログを出力したOS又はAPの名称であり、ログファイル名315はログ出力プログラム313が出力したログファイルの名称である。ログ事象311、ログ事象結果312及びデータ格納クラス314については後述する。マネージャは、マネージャを格納する計算機の名316とホストIPアドレス317を格納する。エージェントは、ログを入力したエージェントを格納する計算機の名318とホストIPアドレス319を格納する。監視対象は、監視対象とするサーバ等の計算機の名320とホストIPアドレス321を格納する。時刻はログ出力時刻322と補正時刻323から成る。ログ出力時刻322はログを出力した計算機の局所の時刻であり、補正時刻323はマネージャ10を格納する計算機の時刻に基づいて補正した時刻である。フィルタリングルール名324は、ログデータを正規化するときに応用したフィルタリングルールの名称である。

【0032】図7は、ユーザ情報クラス302のデータ構成を示す図である。ユーザ情報クラス302は、ログインしたユーザに関する情報を記録するものであり、ユーザ名303、ユーザID (UID) 304、ユーザ変更した後のユーザ名305、変更後のUID306、ユーザのセキュリティレベル307、計算機やファイルへのアクセス308、アクセスした結果309及びユーザが操作した端末装置の名称 (端末名310) を格納する。

【0033】図8は、サービス情報クラス303のデータ構成を示す図である。サービス情報クラス303は、ユーザに提供したサービスについての情報を記録する。サービスは、サービス名340、サービスバージョン341、サービス提供のために起動したプロセスの名称 (プロセス名342) 及びプロセスID343を格納する。

【0034】図9は、アドレス情報クラス304のデータ構成を示す図である。アドレス情報クラス304は、

他計算機とコネクションを行ったときの情報を記録するものであり、接続元及び接続先のホスト名、IPアドレス、MACアドレス、ポート番号の他にコネクション状態、コネクションの開始時刻と終了時刻及び他計算機へのアクセス結果を格納する。

【0035】図10は、ファイル情報クラス305のデータ構成を示す図である。ファイル情報クラス305は、ユーザが作成又は変更したファイルについてファイル名、変更前のアクセス情報及び変更後のアクセス情報を記録する。アクセス情報は、ファイルの作成時刻、最終修正時刻、最終アクセス時刻、ファイルのinode番号、アクセス許可の有無、UID、グループID (GID) 及び最終的なファイルのサイズを格納する。

【0036】図11は、トラフィック情報クラス306のデータ構成を示す図である。トラフィック情報クラス306は、メール管理プログラム、ファイル転送プログラム等が出力するログ情報であり、ネットワークを介するデータやメッセージの受信バイト数、送信バイト数及びデータの転送時間 (処理時間) を記録する。

【0037】図12は、個別情報クラス307のデータ構成を示す図である。個別情報クラス307は、オプションであり、プログラムが出力するメッセージテキストの原文そのままの情報である。

【0038】図13は、データ格納クラス314のデータ構成を示す図である。"T1" はデータ格納クラス314を識別するためのタグであり、"L1" は存在する情報クラスを指定する領域V01の長さを示す。"V1" は、各正規化ログデータ300に含まれる情報クラスを指定する領域であり、情報クラスの指定の開始を示すタグ、情報クラス識別子の長さ及び情報クラス識別子を順番に指定する。情報クラス識別子の長さは可変である。各情報クラスを識別する番号をxとすると、Tx (x ≥ 2) は情報クラスの開始を示すタグであり、Lx (x ≥ 2) はVxの部分の長さであり、Vx (x ≥ 2) は情報クラス識別子である。マネージャ10は、このデータ格納クラス314により各正規化ログデータが有する情報クラスを認識する。

【0039】図14は、正規化項目のうちコード化が可能なもののコード化テーブル600の一例を示す図である。正規化項目のログ事象311が"login"の場合はコード"i"に、ユーザ変更"su"の場合はコード"2"というようにコード化される。コネクト (connect) は、計算機間でファイル転送やプログラム間通信を行う際のコネクションに関するログ事象を示す。ファイルは内容変更されたファイルに関するログ事象、ジョブはジョブの起動/停止、終了状態に関するログ事象である。メールはメール使用に関するログ事象を示す。ログ事象結果312はログ事象の結果であり、成功/失敗かを区分する。ユーザ情報クラス302のアクセス権335についてはあり又はなしを区分する。ユー

情報クラス302及びアドレス情報クラス304のアクセス結果については、成功か失敗かを区別する。ファイル情報クラス305のアクセス許可については、あり又はなしを区別する。

【0040】図15〜図20は、ルールのデータ形式を示す図である。

【0041】図15は、マネージャールール450の一例を示す図である。DB_MAX451は正規化ログデータベース50が正規化ログデータを保存可能な期間を示す保存期間を定義する。RULE_MAX452はルール配布に使用できる最大の通信路数を定義する。LOG_MAX453はログ収集に使用できる最大の通信路数を定義する。ルール配布106やログ収集107は、この多重度数だけ通信路を使用できるが、処理の要求からの多重度より大きい場合は通信路を順番に使用し、通信路が空くまで待った後、残りの処理を実行する。

【0042】図16は、エージェント20の動作条件ルール470の一例を示す図である。MANAGER_ADDRESS471はマネージャのIPアドレスを定義し、FILE_MAXSIZE472は正規化ログファイル40が使用できる最大の記憶容量を定義する。

【0043】図17は、ログファイル監視ルール500の一例を示す図である。TARGET_LOGは監視対象ログファイル名を定義し、FORMATによりファイルの形式（SEQ：シーケンシャル形式、WRAP：ラップアラウンド形式）を定義し、INTERVALにより監視間隔の時間（たとえば、10分間隔）を定義する。FMT_NAMEは当該ログファイルを正規化するときに適用するルールを定義する。例により説明すると、

TARGET_LOG=/usr/adm/syslog.log, FORMAT=SEQ, INTERVAL=10m, FMT_NAME=abc;

は、シーケンシャル形式のファイル/usr/adm/syslog.log、ログを10分間隔で監視し、フォーマットルールabcにより正規化処理を行うことを示す。FMT_NAMEの指定がない場合は、エージェント20の間で共通のフォーマットルールにより正規化を行う。

【0044】図18と図19は、フォーマットルール510及び515の一例を示す図である。ログデータがテキスト形式の場合はFMT_NAMEを適用し、バイナリ形式の場合はFMT_NAMEを適用する。REGEXは“文字列n”は、ログデータを選択する条件を示し、ログデータ中に文字列nが存在すれば、以下に示す規則に従ってログデータを正規化することを示す。&&は論理積（AND）を示し、複数の文字列の存在を選択条件とすることができる。|はthenを意味し、以下ログデータの文字列のシーケンスに従って文字列から順に正規化項目を拾って行くことを意味する。正規化項目とは、各情報クラスで定義されるデータ項目のことである。:

は、ログデータの先頭から順番にポイントを示す正規化項目に対応する値を切り出すための区切り文字である。正規化項目に続いて[]内に指定される文字は、任意のポイントから認識する文字列の長さ、またはその文字列が可変長の場合に認識する最後の文字を指定する。SKIPは、ログデータの先頭から順番にポイントを示すポイントがいった場合、正規化項目に関連ない文字列が存在する場合にその文字列を読み飛ばすことを意味し、[]内には読み飛ばす文字数又は認識する最後の文字となる“区切り文字”を指定する。区切り文字を指定した場合は、その区切り文字まで読み飛ばす。以下フォーマットルールの例を挙げる。

(a) フォーマットルールA

FMT.T.REGEX = "SU" && REGEX = "*" | ログ事象 = "2" | ログ事象結果 = "0" | SKIP[" "]; ログ出力時刻[10]; SKIP[3]; 端末名[" "]; ユーザ名["-"]; 変更後のユーザ名[" "];

(b) フォーマットルールB

FMT.T.REGEX = "connect" && REGEX = "refused" | ログ事象 = "3" | ログ事象結果 = "1" | ログ出力時刻[15]; 接続先ホスト名[" "]; プロセス名["[]"]; プロセスID["[]"]; SKIP["from"]; 接続元ホスト名[" "];

図21は、ログファイルに格納されているメッセージテキストの原文の例を示す図である。メッセージテキスト551及び552は、OSが出力するユーザ変更に關するメッセージテキストの例である。メッセージテキスト553〜556は、OSがコネクション時に出力するメッセージテキストである。メッセージテキスト556及び557は、OSのジョブ管理が出力するメッセージテキストである。

【0045】メッセージテキスト551をフォーマットルールAによって正規化する。

- ・ログ事象311=2(su)
- ・ログ事象結果312=0(成功)
- ・ログ出力時刻322=1/30 11:18のエボクタイム
- ・端末名337=typ5
- ・ユーザ名330=fujino
- ・変更後のユーザ名332=rootとなる。

【0046】メッセージテキスト554をフォーマットルールBによって正規化する。

- ・ログ事象311=3(connect)
- ・ログ事象結果312=1(失敗)
- ・ログ出力時刻322=Jan 12 13:12:15のエボクタイム
- ・接続先ホスト名354=hosta
- ・接続先IPアドレス355=hostaをIPアドレスに変更した値

・プロセス名 342=ipd
 ・プロセス ID 343=1111
 ・接続元ホスト名 350=hostb
 ・接続元 IP アドレス 351=hostb を IP アドレスに変換した値
 となる。

【0047】図20は、フィルタリングルール520の一例を示す図であり、ログファイル管理204が該当する正規化ログデータだけを格納するためのルールである。FILTはフィルタリングルールであることを示し、正規化項目が指定した文字列やコードであったり、指定した時間帯の正規化ログデータであった場合、そのような条件に適合する正規化ログデータだけを抽出して格納する。==はイコールを、!=はnotイコールを、&&は論理積ANDを、||は論理和ORを、-は時間間隔をそれぞれ意味する。

【0048】図22は、正規化ログデータベース50のデータ構造を示す図である。正規化ログデータベース50は、正規化ログデータを補正時刻630の順に配列して格納する。ある補正時刻630から共通情報クラス631のログデータにチェインする。また共通情報クラス631からこれに続いて存在する情報クラスのログデータに次々とチェインする。情報クラスに対応して示される検索キーは、正規化ログデータを検索するときキーとして使用される正規化項目を示すものである。また補正時刻630から次の補正時刻630へチェインが張られている。ユーザは、補正時刻630によって、また該当する情報クラスの情報キーを指定することによって正規化ログデータベース50から目的とする正規化ログデータを効率的に抽出することができる。

【0049】図23は、マネージャ10のルール配布106の処理の流れを示すPAD図である。ルール配布106は、初期設定(ステップ701)後、プロセス管理102から終了要求が来るまでループし(ステップ702)、イベントを持つ(ステップ703)。イベントには、スケジューラ101からのルール配布要求(ステップ704)とプロセス管理102からの終了要求(ステップ712)がある。

【0050】ルール配布要求(ステップ704)を受信した場合は、ルール管理105を介して配布するルールと配布先であるエージェント20の一覧を取得し(ステップ705)、構成管理104からエージェント20の動作状態及び応答時間の情報を取得する(ステップ706)。取得したエージェント20のping応答時間を応答時間の小さい順番に並べ替える(ステップ707)。このとき動作していないエージェント20については、応答時間を無限大と解釈する。ルールを配布するエージェント20の数だけループし(ステップ708)、エージェント20が動作している場合(ステップ709YES)は応答時間の小さい順番にルールを配布

し(ステップ710)、エージェント20が動作していない場合はルール配布失敗のメッセージを出力する(ステップ711)。ルール配布に当たっては、RULE_MAX_MAX452を適用する。配布したルールは、エージェント20のルール管理203に転送される。

【0051】終了要求を受信した場合は(ステップ712)、ループを抜けて終了処理を行う(ステップ713)。

【0052】図24は、マネージャ10のログ収集107の処理の流れを示すPAD図である。ログ収集107は、初期設定(ステップ801)後、プロセス管理102から終了要求が来るまでループし(ステップ802)、イベントを持つ(ステップ803)。イベントには、スケジューラ101からのログ収集要求(ステップ804)、エージェント20からの起動通知(ステップ811)、及びプロセス管理102からの終了要求(ステップ816)がある。

【0053】ログ収集要求を受信した場合は(ステップ804)、スケジューラ101からログを収集するエージェント20の一覧を取得し(ステップ805)、構成管理104からこれらのエージェント20の動作状態及び応答時間の情報を取得するとともに、応答時間の高い順番にエージェント20をソートする(ステップ806)。ログを収集するエージェント20の数だけループし(ステップ807)、エージェントが動作している場合(ステップ808YES)はログを収集し(ステップ809)、エージェントが動作していない場合(ステップ808NO)はログ収集失敗メッセージを出力する(ステップ810)。ログ収集に当たっては、LOG_MAX453を適用する。ログ収集107は、エージェント20のログファイル管理204を介して正規化ログファイル40を収集する。

【0054】エージェント起動通知を受信した場合は(ステップ811)、構成管理104からエージェント20とのping(ICMPエコーリクエスト)の応答時間を取得する(ステップ812)。ICMP(Internet Control Message Protocol)は、通信ネットワークの管理に関する国際的な標準規格の一つであるアイ・エイ・ビー(Internet Activities Board)の管理標準である。ICMPを使用すると、IPノード(例えばコンピュータ)が他のIPノードと通信可能であるか否かを確認できる。またpingを使用すると、任意のIPノードと通信可能であるか否かの動作状態と応答時間を取得できる。ping応答時間を取得できた場合(ステップ813YES)は、マネージャ10の現在時刻にこの応答時間から得られる通信時間を加算した時刻を起動通知を発行したエージェント20へ通知する(ステップ814)。マネージャの時刻をマネージャから伝えるためには、マネージャの時刻にマネージャから

エージェントへの通信時間を加えた時刻を通知すればよい。pingの応答時間は、マネージャからエージェントとエージェントからマネージャ、つまり行きと返りの通信時間を加えた時間間隔である。そこでping応答時間の1/2を通信時間として利用する。すなわちマネージャは、次の計算式によりエージェントの時刻を推定しエージェントへ通知する。

(エージェントの時刻) = (マネージャの時刻) + (ping応答時間) / 2

応答時間を取得できなかった場合(ステップ813N O)、すなわちネットワーク管理システム60から情報を得られない場合は、単にマネージャ10の現在時刻をエージェント20へ通知する(ステップ815)。エージェント20は、マネージャ10の現在時刻を取得して正規化ログデータの補正時刻323に適用する。

【0055】終了要求を受信した場合は(ステップ816)、ループを抜けて終了処理を行う(ステップ817)。

【0056】図25は、マネージャ10の構成管理104の処理の流れを示すPAD図である。構成管理104は、初期設定(ステップ901)後、プロセス管理102から終了要求が来るまでループし(ステップ902)、イベントを持つ(ステップ903)。イベントにはルール配布106からのエージェント情報格納要求(ステップ904)、ルール配布106やログ収集107からのエージェント情報取得要求(ステップ909)と、プロセス管理102からの終了要求(ステップ913)がある。

【0057】エージェント情報格納要求を受信した場合は(ステップ904)、ルール配布106からルールを配布したエージェント20の情報を取得する(ステップ905)。エージェント20の情報とは、ルールの配布時刻、配布したルール名などである。ネットワーク管理システム60と通信可能であるとき(ステップ906YES)は、取得したエージェント情報をネットワーク管理システム60に渡す(ステップ907)。ネットワーク管理システム60は、受信したルール配布の履歴情報をネットワーク管理のために利用可能である。ネットワーク管理システム60と通信できないとき(ステップ906NO)は、エージェント情報を構成管理104が保存するファイルへ格納する(ステップ908)。

【0058】エージェント情報取得要求を受信した場合は(ステップ909)、ネットワーク管理システム60と通信できるとき(ステップ910YES)には、ネットワーク管理システム60からエージェント20の一覧、動作有無やpingの応答時間等を取得しこれらの情報を要求元に送す(ステップ911)。ネットワーク管理システム60と通信できないとき(ステップ910NO)には、構成管理104のファイルからエージェントの一覧についての情報を取得して要求元に送す(ス

ップ912)。

【0059】終了要求を受信した場合は(ステップ913)、ループを抜けて終了処理(ステップ914)を行う。

【0060】図26は、マネージャ10のデータベース管理109の処理の流れを示すPAD図である。データベース管理109は、初期設定(ステップ1001)し、正規化ログデータベース50に格納している正規化ログデータの保存期間の確認を要求(ステップ1002)した後、プロセス管理102から終了要求が来るまでループし(ステップ1003)、イベントを持つ(ステップ1004)。イベントには、ログ収集107からの正規化ログデータ格納通知(ステップ1005)、ログ解析103からの正規化ログデータ抽出要求(ステップ1008)、データベース管理109自身が正規化ログデータの保存期間を確認するための要求(ステップ1010)と、プロセス管理102からの終了要求(ステップ1014)がある。

【0061】正規化ログデータ格納通知を受けた場合は(ステップ1005)、ログ収集107から正規化ログデータを取得し正規化ログデータベース50に格納する(ステップ1006)。格納に当たっては、図22のデータ構造に従って正規化ログデータを格納する。正規化ログデータは補正時刻630の順に配列されるので、この順に従って正規化ログデータをマージする。また正規化ログデータの保存期間確認を要求する(ステップ1007)。

【0062】正規化ログデータ抽出要求を受信した場合は(ステップ1008)、指定された検索キーにより正規化ログデータベース50を検索し、その結果抽出したデータを要求元へ応答する(ステップ1009)。

【0063】保存期間確認要求(ステップ1010)を受信した場合は、正規化ログデータベース50に格納されている正規化ログデータの一番古い補正時刻630と現在時刻の差と、DB_MAX451とを比較し(ステップ1011)、保存期間より古い正規化ログデータを保存しているときは古い正規化ログデータを削除し(ステップ1012)、運用者に知らせるために削除メッセージを出力する(ステップ1013)。

【0064】終了要求を受信した場合は(ステップ1014)、ループを抜けて終了処理(ステップ1015)を行う。

【0065】図27は、エージェント20のログ入力202の処理の流れを示すPAD図である。ログ入力202は、初期設定し(ステップ1101)、エージェント20のルール管理203からログファイル監視ルール50とフォーマットルール510、515等を取得(ステップ1102)後、プロセス管理201から終了要求が来るまでループし(ステップ1103)、イベントを持つ(ステップ1104)。イベントには、ログファイ

ル管理204からのログ入力中断要求(ステップ1105)とログ入力再開要求(ステップ1107)、プロセス管理201からの終了要求(ステップ1109)及び時間監視によるタイム割り込みがある。

【0066】ログ入力中断要求を受信した場合は(ステップ1105)、ログ入力を中断する(ステップ1106)。中断する要因は、正規化ログファイル40の容量がFILE_MAXSIZE47に達したときである。

【0067】ログ入力再開要求を受信した場合は(ステップ1107)、ログ入力を再開する(ステップ1108)。再開する要因は、正規化ログファイルをマネージャ10へ転送したときである。

【0068】終了要求を受信した場合は(ステップ1109)、ループを抜けて終了処理(ステップ1117)を行う。

【0069】ログファイル監視ルール500に設定された監視間隔に従ってログファイルの監視時刻になったとき、監視対象ログファイル置30をオープンし(ステップ1111)、このログファイルのファイル管理情報を取得する。ファイル管理情報が前回取得したものと同じか否かを確認する(ステップ1112)。同じ場合は(ステップ1112YES)、前回オープンしたファイルと同じ内容であるため前回のファイルのオフセットからログデータを入手する(ステップ1113)。前回のファイルのオフセットは、当該ファイルについて前回入力済のレコードの次のレコードを指している。異なる場合は(ステップ1112NO)、新しいファイル(前回オープンしたファイルは削除された等)であると解釈し、先頭からログデータを入手する(ステップ1114)。その後、入力したログデータを正規化し(ステップ1115)、正規化ログデータをログファイル管理204へ通知する(ステップ1116)。ログデータの正規化は、上記のようにフォーマットルール510、515等を適用して行う。正規化ログデータをログファイル管理204に渡した後、当該ログファイルをクローズし、監視間隔に従って次の監視時刻にタイムを設定する。

【0070】図28は、エージェント20のログファイル管理204の処理の流れを示すA/D図である。ログファイル管理204は、初期設定し(ステップ1201)、ルール管理203から動作条件ルール470とフィルタリングルール520を取得(ステップ1202)した後、プロセス管理201から終了要求が来るまでループし(ステップ1203)、イベントを持つ(ステップ1204)。イベントには、ログ入力202からの正規化ログデータ格納通知(ステップ1205)、マネージャ10のログ収集107からのログ収集要求(ステップ1209)、ログファイル管理204自身からの正規化ログファイル容量確認要求(ステップ1211)、マネージャ10のログ収集107からのマネージャ時刻の

通知(ステップ1216)と、プロセス管理201からの終了要求(ステップ1218)がある。

【0071】正規化ログデータ格納通知を受信した場合は(ステップ1205)、ログ入力202から正規化ログデータを取得し(ステップ1206)、エージェントとマネージャの時間差(ステップ1217の処理結果)とログ出力時刻322から補正時刻323を計算する。ログ入力202から取得した正規化ログデータはこの補正時刻323を追加して正規化ログファイル40に格納する(ステップ1207)。取得した正規化ログデータにフィルタリングルール520を適用して条件に合致する正規化ログデータのみを正規化ログファイル40に格納する。次に正規化ログファイル容量確認要求を発行する(ステップ1208)。

【0072】ログ収集要求を受信した場合は(ステップ1209)、正規化ログファイル40中の正規化ログデータを補正時刻323の順にソートした後、MANAGER_ADDRESS471に示されるマネージャ10へ転送する(ステップ1210)。

【0073】正規化ログファイル容量確認要求を受信した場合は(ステップ1211)、正規化ログファイルの使用容量とFILE_MAXSIZE47を比較し(ステップ1212)、最大サイズに達したとき(ステップ1212YES)は、ログ入力202へ中断通知を発行する(ステップ1213)。最大サイズに達していないとき(ステップ1212NO)は、前回中断要求を発行したか確認し(ステップ1214)、発行していたとき(ステップ1214YES)は、ログ入力202へログ入力再開要求を通知する(ステップ1215)。

【0074】マネージャ時刻の通知を受信した場合は(ステップ1216)、エージェントとマネージャのコンピュータ時刻の差を計算する(ステップ1217)。

【0075】終了要求を受信した場合は(ステップ1218)、ループを抜けて終了処理(ステップ1219)を行う。

【0076】

【発明の結果】以上説明したように本発明によれば、エージェントが複数のログファイルを監視し得る形式で出力されたログデータを入力した後、正規化を行い共通的なデータ形式に変換する。また必要なログデータだけを抽出し、ログデータの出力時刻としてマネージャの時刻に合わせた補正時刻を使用するようにしたので、適用者はネットワークに存在する複数のコンピュータのログデータを統一したデータ形式及び時刻に基づいて解析することができる。

【0077】またマネージャが格納するログデータについては、所定期間のログデータを保存するようにしたので、古いログデータから順に削除する形でログ情報の脱報を抑制できる。

【0078】さらに収集したログデータを補正時刻及び

正規化項目によって検索可能としたので、運用者は必要なログ情報を容易に取得できる。

【図面の簡単な説明】

【図1】実施形態のネットワークシステムの構成図である。

【図2】実施形態のマネージャとエージェントが行う処理動作の概略を示す図である。

【図3】実施形態のマネージャ10の構成を示す機能ブロック図である。

【図4】実施形態のエージェント20の構成を示す機能ブロック図である。

【図5】実施形態の正規化ログデータの概略構成を示す図である。

【図6】実施形態の基番情報クラスのデータ構成を示す図である。

【図7】実施形態のユーザ情報クラスのデータ構成を示す図である。

【図8】実施形態のサービス情報クラスのデータ構成を示す図である。

【図9】実施形態のアドレス情報クラスのデータ構成を示す図である。

【図10】実施形態のファイル情報クラスのデータ構成を示す図である。

【図11】実施形態のトラフィック情報クラスのデータ構成を示す図である。

【図12】実施形態の個別情報クラスのデータ構成を示す図である。

【図13】実施形態のデータ格納クラスのデータ構成を示す図である。

【図14】正規化項目のコード化テーブルの例を示す図である。

【図15】マネージャルールの例を示す図である。

【図16】エージェントの動作条件ルールの例を示す図

である。

【図17】エージェントのログファイル監視ルールの例を示す図である。

【図18】エージェントのフォーマットルール（その1）の例を示す図である。

【図19】エージェントのフォーマットルール（その2）の例を示す図である。

【図20】エージェントのフィルタリングルールの例を示す図である。

【図21】監視対象ログファイルのログデータであるメッセージリストの例を示す図である。

【図22】実施形態の正規化ログデータベースのデータ構造を示す図である。

【図23】実施形態のマネージャが行うルール配布の処理の流れを示すPAD図である。

【図24】実施形態のマネージャが行うログ収集の処理の流れを示すPAD図である。

【図25】実施形態のマネージャが行う構成管理の処理の流れを示すPAD図である。

【図26】実施形態のマネージャが行うデータベース管理の処理の流れを示すPAD図である。

【図27】実施形態のエージェントが行うログ入力の処理の流れを示すPAD図である。

【図28】実施形態のエージェントが行うログファイル管理の処理の流れを示すPAD図である。

【符号の説明】

10・・・マネージャ、20・・・エージェント、40・・・正規化ログファイル、50・・・正規化ログデータベース、60・・・ネットワーク管理システム、104・・・構成管理、106・・・ルール配布、107・・・ログ収集、109・・・データベース管理、110・・・ルール、202・・・ログ入力、204・・・ログファイル管理、205・・・ルール

【図7】

図7

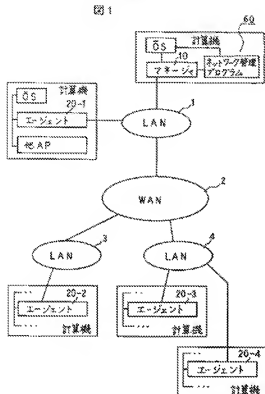
ユーザ情報クラス ~302	
ユーザ名	~330
UID	~331
変更後のユーザ名	~332
変更後のUID	~333
セキュリティレベル	~334
アクセス権	~335
アクセス結果	~336
端末名	~337

【図8】

図8

サービス情報クラス ~303	
サービス名	~340
サービスバージョン	~341
プロセス名	~342
プロセスID	~343

【図1】



【図5】

図5

正規化ログデータ	
共通情報クラス	～301
ユーザ情報クラス	～302
サービス情報クラス	～303
アドレス情報クラス	～304
ファイル情報クラス	～305
・	
・	
トラフィック情報クラス	～306
個別情報クラス	～307

【図6】

図6

共通情報クラス		～301
正規化バージョン		～310
ログ種別	ログ事象	～311
	ログ事象結果	～312
	ログ出力プログラム	～313
	データ格納クラス	～314
	ログファイル名	～315
マネージャ	ホスト名	～316
	ホストIPアドレス	～317
エージェント	ホスト名	～318
	ホストIPアドレス	～319
監視対象	ホスト名	～320
	ホストIPアドレス	～321
時刻	ログ出力時刻	～322
	補正時刻	～323
フィルタリングルール名		～324

【図11】

図11

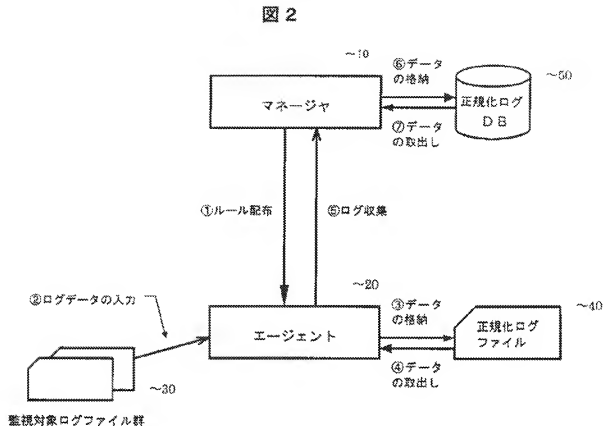
トラフィック情報クラス	～306
受信バイト数	
送信バイト数	
処理時間	

【図12】

図12

個別情報クラス	～307
メッセージテキスト	

【図2】



【図9】

図9

アドレス情報クラス		～304
接続元ホスト名	～350	
接続元IPアドレス	～351	
接続元MACアドレス	～352	
接続元ポート番号	～353	
接続先ホスト名	～354	
接続先IPアドレス	～355	
接続先MACアドレス	～356	
接続先ポート番号	～357	
コネクション状態	～358	
開始時刻	～359	
終了時刻	～360	
アクセス結果	～361	

【図15】

図15

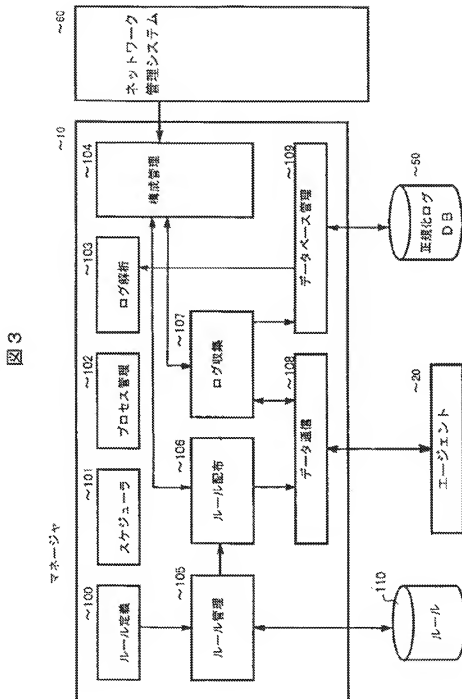
マネージャルール		～400
DB_MAX: 正規化ログデータの保存期間;	～401	
RULE_MAX: ルール配布の最大歩数;	～402	
LOG_MAX: ログ収集の最大歩数;	～403	

【図16】

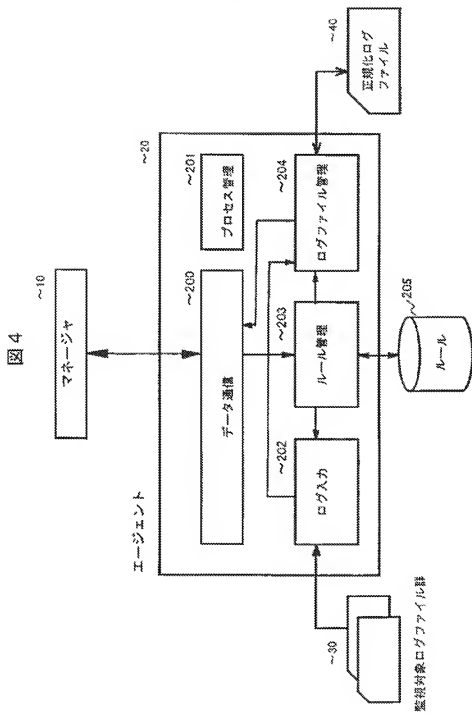
図16

動作条件ルール		～410
BANNER_ADDRESS: マネージャのIPアドレス;	～411	
FILE_MAXSIZE: 正規化ファイルの最大サイズ;	～412	

【図3】



【図4】



【図10】

図10

ファイル情報クラス

～305

ファイル名	
変更後	作成時刻
	最終修正時刻
	最終アクセス時刻
	i-node番号
	アクセス許可
	UID
	GID
	サイズ
変更前	作成時刻
	最終修正時刻
	最終アクセス時刻
	i-node番号
	アクセス許可
	UID
	GID
	サイズ

【図13】

図13

データ格納クラス

～314

T1	L1	V1 (存在する情報クラスの指定領域)					
		T2	L2	V2	T5	L5	V5
						

【図18】

図18

フォーマットルール (その1)

～510

```
FWT_T: REGTEXT == "文字列1" | ログ事象 == "T" | ログ事象結果 == "R"
| 正規化項目1 | 文字列2 | 正規化項目2 | 文字列3 | ...
FWT_T: REGTEXT == "文字列2" && REGTEXT == "文字列3" | ログ事象 == "R"
| ログ事象結果 == "T" | 正規化項目1 | 文字列4 | REGID | 読み取り文字列 |
| 正規化項目4 | 文字列5 | ...
```

【図19】

図19

フォーマットルール (その2)

～510

```
FWT_B: REGTEXT == "文字列5" && REGTEXT == "文字列6" | ログ事象 == "T"
| ログ事象結果 == "T" | 正規化項目1 | ログ文字列 | REGID | 読み取り文字列 | 正規
化項目2 | 文字列7 | ...
```

【図17】

図17

ログファイル管理ルール

～500

```
FARBIT_LOG: "監視対象ログファイル名1", FPMBA1-REG, INTERVAL-時間,
FWT_NAME/フォーマットルール名;
FARBIT_LOG: "監視対象ログファイル名2", FPMBA2-REG, INTERVAL-時間;
```

【図21】

図21

メッセージデキスト

～550

```
201 01/20 11:10 + httpd fujimoriout ~551
201 01/21 11:52 + httpd morita-abc ~552
201 4 12:30 10 hosts fpt[111]: connect from 178.213.252.12 ~553
201 12 12:12 15 hosts fpt[111]: refused connect from hosts ~554
201 18 15:10 50 hosts fpt[111]: connect to hosts ~555
201 25 10:12:34 hosts job[222]: ジョブABCを開始しました。 ~556
201 25 10:10:10 hosts job[222]: ジョブABCが異常終了しました。 ~557
```

【図20】

図20

フォーマットルール

～570

```
FPT: 正規化項目1 == "文字列1";
FPT: 正規化項目2 == "文字列2" | 正規化項目3 == "文字列3";
FPT: 正規化項目4 == "文字列4" && 正規化項目5 == "文字列5";
FPT: 正規化項目6 == "時刻1" ~ "時刻2" && 正規化項目7 == "文字列7";
```


【図14】

図14

正規化項目のコード化テーブル

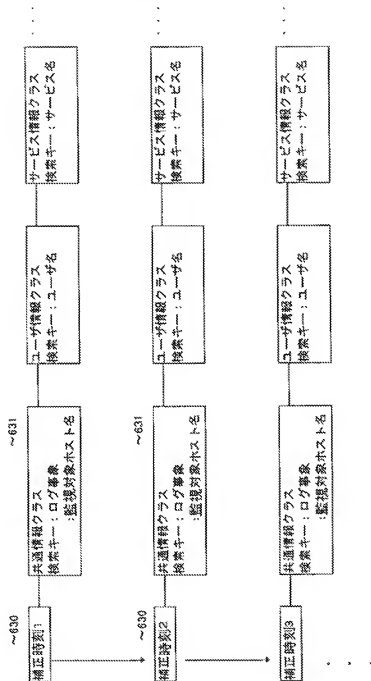
～590

正規化項目	内容	コード
ログ事象	login	1
	su (ユーザ変更)	2
	connect	3
	ファイル	4
	ジョブ	5
	メール	6
ログ事象結果	成功	0
	失敗	1
アクセス権	あり	0
	なし	1
アクセス結果	成功	0
	失敗	1
アクセス許可	あり	0
	なし	1

【図22】

図 22

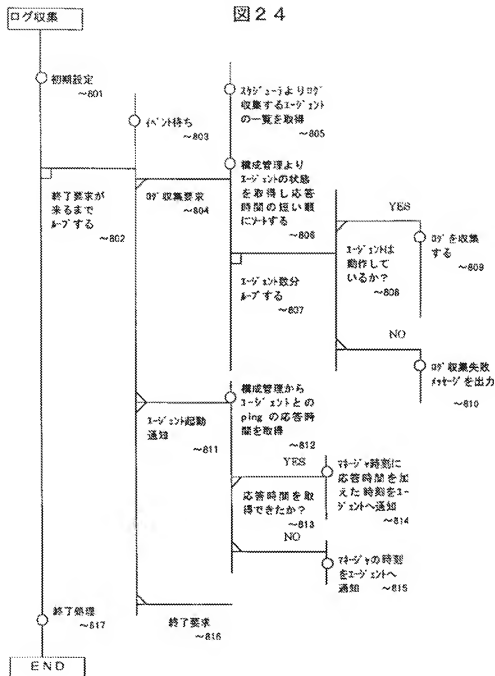
5.0: 正規化ログデータベース



23

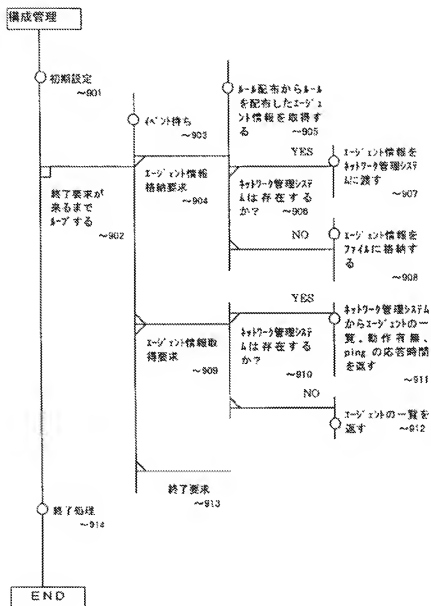


【例2-4】



【図25】

図25

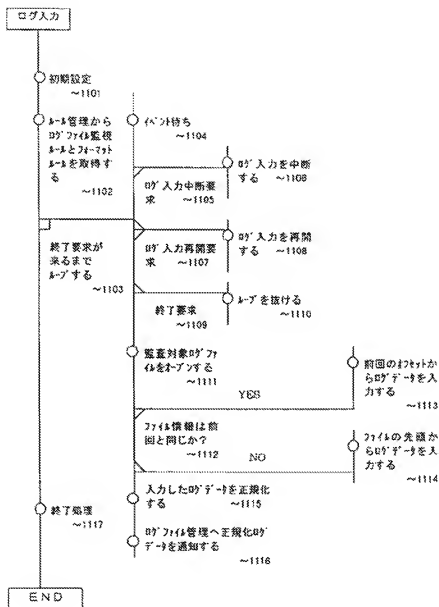


26



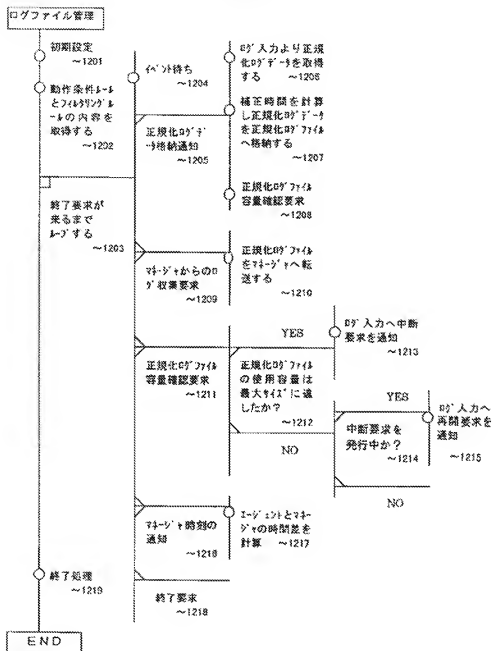
【図27】

図 27



【図28】

図28



フロントページの続き

(72)発明者 中野 武紀
 東京都千代田区大手町二丁目6番2号 株式会社日立情報ネットワーク内

(72)発明者 森田 眞司
 神奈川県横浜市中区戸塚町5030番地 株式会社日立製作所ソフトウェア開発本部内

(72)発明者 山田 寛
神奈川県横浜市中区戸塚町5030番地 株
式会社日立製作所ソフトウェア開発本部内

(72)発明者 新村 美貴
神奈川県横浜市中区戸塚町5030番地 株
式会社日立製作所ソフトウェア開発本部内